

周报总结

2015-08-23

本周总结:

工作部分:

上周巫老师走之前提到了一种 K-means 的 Evolutionary Clustering 版本, 这周开始的时候没我主要是找了一些 Evolutionary Clustering 算法相关资料。

搜集资料之后, 从中找了一篇比较适合我们当前情况的论文, 其中 "Evolutionary Clustering"(written by Deepayan Chakrabarti、Ravi Kumar、Andrew Tomkins)。选择这个文章的原因有以下几点:

1 在文章中作者主要对两个算法进行了改进, 一个是 kmeans 另一个是层次聚类, 恰好是之前使用过的方法。

2 Evolutionary Clustering 算法考虑到了时间序列模型中不同时间前后对聚类的影响, 不但让聚类结果更加准确, 也更加符合真实时间序列模型本身。

3 算法针对时间序列的数据进行优化, 很多细节符合我们网易数据的形式。比如说, 网易数据有一些玩家是中途出现的, 有一些玩家中间时间点可能不会上线。这些形式的数据在本文中都有所考虑到。因此很符合我们的实际情况。

本周主要是找资料多一些, 太久没有看论文, 感觉看起来有吃力。下周希望能把论文中的应用进行一下实现, 看下聚类效果是否有显著的提升。

下周工作:

实现论文算法。同时希望能尽快吧当前论文结掉, 开始自己真正感兴趣的研究方向。